Caractérisation statique et dynamique des voyelles dans des séquences VV

Julien Millasseau, Olivier Crouzet

Caractérisation statique et dynamique des voyelles dans des séquences VV.

Julien Millasseau, Olivier Crouzet
Laboratoire de Linguistique de Nantes — LLING — UMR 6310 CNRS / Université de Nantes
Chemin de la Censive du Tertre, 44300 Nantes, France
julien.millasseau@etu.univ-nantes.fr, olivier.crouzet@univ-nantes.fr

RÉSUMÉ

Nous étudions les indices acoustiques liés à la caractérisation statique et / ou dynamique des voyelles du français. Nous avons analysé les caractéristiques formantiques de six réalisations vocaliques ainsi que les transitions formantiques de seize combinaisons V_1V_2 impliquant ces 6 voyelles afin d'évaluer les contributions des indices dynamiques liés aux transitions entre voyelles et des indices statiques de fréquence. Les mesures correspondantes sont issues d'un protocole dans lequel le débit de parole était influencé expérimentalement afin de provoquer d'éventuelles variations de vitesse de transition. Les résultats ne permettent pas de départager ces deux hypothèses mais montrent que les indices dynamiques pourraient être aussi fiables que les mesures statiques. Des pistes d'extension de ce travail sont proposées qui pourraient contribuer de manière plus informative à cette problématique.

ABSTRACT

Static and dynamic characterization of vowels in VV sequences.

The present study aims at evaluating the respective contributions of static and dynamic cues to vowel classification. Formant cues from six french vowels and sixteen V_1V_2 slope transitions were extracted in order to investigate the respective contributions of dynamic and static cues that would be respectively related to transitions or center frequencies. The corresponding data were collected from a dedicated task in which speech rate was influenced experimentally in order to trigger potential variations of rate of change within the transitions. The current results do not favour any of the two potential accounts but show that dynamic cues may be as reliable as static ones. Follow-ups to this protocol are offered that may contribute to this issue more informatively.

MOTS-CLÉS : Dynamique de la parole, Analyse Discriminante Linéaire, Classification des voyelles.

KEYWORDS: Speech Dynamics, Linear Discriminant Analysis, Vowel classification.

1 Introduction

Les études menées en perception et en production de la parole ont montré une grande variabilité articulatoire et acoustique des voyelles (Peterson & Barney, 1952). Ceci est dû à la fois à la diversité des locuteurs, aux variations de débit mais également aux contextes et même aux variations internes à une langue. Certaines études ont soutenu que ces variations étaient mues par la dynamique contextuelle (Lindblom & Studdert-Kennedy, 1967; Strange *et al.*, 1979), et qu'une perspective statique ne pouvait donc suffire à caractériser les voyelles, aussi bien en perception qu'en production. D'autres travaux ont

étendu cette hypothèse aux transitions de séquences CV, VC, CVC (Strange *et al.*, 1983; Strange, 1989; Nearey, 1989; Andruski & Nearey, 1992; Jenkins *et al.*, 1999; Hillenbrand *et al.*, 2001).

Les tâches de perception menées par Strange (Strange *et al.*, 1983; Strange, 1989) et plus tard Jenkins (Jenkins *et al.*, 1999) sur les syllabes dites *silent-center* ont montré qu'une voyelle remplacée par un silence d'une durée équivalente dans une séquence CVC reste aisément identifiable. Selon Strange (Strange *et al.*, 1983; Strange, 1989), ce sont les transitions *on glide* et *off glide* qui permettent, malgré « l'absence » de voyelle, de percevoir ses propriétés linguistiques. Dans cette même optique, Hillenbrand à montré que les éléments les plus pertinents dans une tâche de classification des voyelles sont les valeurs de f_0, la durée ainsi que les valeurs de fréquence de $F1$, $F2$ et $F3$ mesurées dans les portions *onset* (à 25% de la voyelle) et *offset* (à 75%). De ces études ressort la nécessité d'une prise en compte de la composante dynamique dans les tâches de perception et de production de la parole, laquelle semble confirmée par des études récentes mettant en avant la relative stabilité des transitions CV (Hillenbrand *et al.*, 2001) et V_1V_2 (Carré, 2009; Divenyi, 2009) produites par différents locuteurs.

D'après les données de Carré (2009), les pentes maximales des transitions V_1V_2 variraient très peu en fonction des locuteurs et ce même lorsque le débit change. Sur la base des données issues d'une expérience dans laquelle les locuteurs devaient produire une série de séquences V_1V_2 ([aV], dans lesquelles V_1 était systématiquement [a] et V_2 une des 9 voyelles orales du français) à 2 débits différents (« normal »vs. « rapide »), Carré (2009) a comparé les taux de variation des fréquences formantiques mesurées au milieu de chaque V_2 et des pentes maximales mesurées à la transition entre les deux voyelles. Il ressort de la représentation en espace $F1 \sim F2$ que les taux de transition formantique des séquences [aV] seraient soumis à une moins grande variation que les fréquences statiques des $F1$ et $F2$ associés à chaque voyelle. Ces constatations sont fondées sur l'observation des mesures de dispersion des données acoustiques recueillies. Selon Carré (2009), ces indices de pente formantique seraient nécessaires à la caractérisation des voyelles et surtout suffisants pour permettre leur identification malgré les effets de la coarticulation et des variations de débit. Cette hypothèse est soutenue par Divenyi (2009) dans une étude perceptive dans laquelle la vélocité des transitions de séquences V_1V_2 artificielles ressort comme l'une des composantes de la réponse perceptive.

Dans cette étude, nous évaluons la pertinence des indices statiques ($F1$, $F2$, $F3$ et f_0) vs. dynamiques (les pentes « maximales » des transitions) pour la classification vocalique. Cette évaluation recourt à des analyses discriminantes linéaires (LDA pour *Linear Discriminant Analysis*), une méthode statistique de classification basée sur l'utilisation de prédicteurs continus pouvant contribuer à la classification en catégories pré-établies. Ce type d'analyse est principalement exploratoire, au sens où il ne nous indique pas si les indices sont cognitivement pertinents pour un locuteur humain, mais il permet d'analyser la contribution statistique des différents prédicteurs à la séparation des classes impliquées du point de vue de la relation acoustique / linguistique. Nous nous sommes également intéressés au rôle du débit dans la composante dynamique. En effet, selon Carré (2009) le débit n'a pas, ou peu, d'impact sur les pentes des transitions. Or nous pourrions nous attendre à ce que celles-ci s'abaissent à un débit lent et s'accroissent à un débit rapide (O'Shaughnessy, 1986).

2 Méthode

Nous avons donc conçu un protocole de production de la parole nous permettant d'analyser les propriétés acoustiques de différentes séquences de voyelles produites par des locuteurs francophones lorsque le débit varie expérimentalement et d'étudier la contribution de différentes informations

acoustiques à la classification statistique des catégories impliquées. Par manque de place, les effets acoustiques des variations de débit ne seront pas présentés dans cet article.

2.1 Participants

Cinq locuteurs francophones (3 femmes et 2 hommes) âgés de 20 à 25 ans ont participé à cette expérience.

2.2 Matériel

Le corpus de séquences consiste en une série de phrases simples construites autour d'une séquence Voyelle-Voyelle. Nous avons ainsi sélectionné 16 paires V_1V_2 appariées 2 à 2 par leur forme symétrique (/ie/, /ei/, /ia/, /ai/, /iu/, /ui/, /iy/, /yi/, /ea/, /ae/, /eo/, /oe/, /ou/, /uo/, /uy/, /yu/) à partir de 6 voyelles orales du français (/i, e, a, o, u, y/). Les paires ont été construites autour d'une variation de un à trois traits distinctifs (cf. Table 1 ; p. ex. /ie/ : [± haut] ; /iu/ : [± arrière / ± arrondi]). Les seules paires impliquant un changement de 3 traits sont la paire /ia/ et son symétrique /ai/.

	− *arrière* − *arrondi*	− *arrière* + *arrondi*	+ *arrière* − *arrondi*	+ *arrière* + *arrondi*
+ *haut* / − *bas*	*i*	*y*		*u*
− *haut* / − *bas*	*e*			*o*
− *haut* / + *bas*			*a*	

TABLE 1 – Représentation en traits des voyelles utilisées dans l'expérience (inspiré de Durand, 2005).

Ces paires ont servi de base à la sélection de séquences de deux mots par recherche automatisée dans la base de données BRULEX (Content *et al.*, 1990). Le premier mot (toujours un nom commun composé de 2 syllabes) se termine par V_1 alors que le second (toujours un adjectif composé de 3 syllabes) commence par V_2. Par exemple, pour la séquence /eo/ nous avons sélectionné les mots « abbé » et « autonome » (/abeotonom/). Ces suites de deux mots ont ensuite été insérées dans la séquence porteuse « *Regarde [ce|cette|ces] MOT1 MOT2* » (p. ex. « *Regarde cet abbé autonome.* ») afin d'être lues par des locuteurs naïfs. Du point de vue du contexte phonétique, V_1 (resp. V_2) est toujours précédée par (resp. suivie de) une occlusive afin de fournir des marqueurs acoustiques pour la procédure de segmentation. Le voisement et le lieu d'articulation de ces occlusives varient de manière non-contrôlée en fonction des mots. Pour chacune des 16 paires V_1V_2, deux suites nom–adjectif différentes ont été sélectionnées pour aboutir à une liste de 32 séquences Mot1–Mot2.

2.3 Procédure

Les locuteurs devaient lire à haute voix les phrases porteuses présentées aléatoirement sur un écran d'ordinateur par un programme en langage `Python` [1] utilisant la librairie `Pygame` [2]. Chaque locuteur était enregistré en 3 blocs successifs correspondant à 3 débits de parole « cible » (rapide, moyen, lent

1. http ://www.python.org
2. http ://www.pygame.org

— toujours dans cet ordre). Le programme consistait en une barre de progression se développant sous la phrase porteuse affichée, à la manière d'un karaoké. La vitesse de progression de cette barre (elle-même corrélée à l'intervalle inter-stimulus et donc au rythme de succession des phrases) représentait le débit cible vers lequel le locuteur devait tendre, le but étant d'influencer le débit effectif de parole. Lorsque la barre de progression atteignait la fin de la phrase orthographiée, l'écran s'effaçait et une nouvelle phrase apparaîssait après un intervalle de 500ms. Les débits cible « attendus » sont exprimés en ms / syllabe. Les 3 débits cible retenus sont 70ms / syllabe, 140ms / syllabe et 190ms / syllabe. Toutes les phrases sont considérées comme étant composées de 8 syllabes (voyelles non-élidées). Les débits impliqués correspondent donc respectivement aux valeurs temporelles suivantes pour les débits rapide (70ms / syll., durée totale de la progression 560ms, ISI [3] 1060ms), moyen (140ms / syll., progression 1120ms, ISI 1620ms) et lent (190ms / syll., progression 1520ms, ISI 2020ms). Avant chaque bloc, les locuteurs étaient soumis à une phase d'entraînement afin de se familiariser avec l'interface et de s'adapter au débit cible. Pendant la passation, les locuteurs avaient à tout moment la possibilité de mettre le programme en pause. Les enregistrements ont été réalisés dans une pièce silencieuse, à l'aide d'un microphone Røde S1 et d'un enregistreur TASCAM DR-680 (format monophonique digitalisé à 44100Hz et encodé sur 16bits dans un fichier WAV).

Les 1440 réalisations issues des différents enregistrements ont été segmentées et transcrites manuellement à l'aide du logiciel Praat (Boersma & Weening, 2014). Deux tires de segmentation / transcription ont été générées : la première correspond à la séquence V_1V_2 dans son ensemble et est utilisée pour l'analyse « dynamique ». La seconde sert à délimiter les zones stables de chacune des deux voyelles V_1 et V_2 pour l'analyse dite « statique ». Les données acoustiques (f_0 et fréquence des trois premiers formants) associées aux transcriptions indiquées dans le deux tires (classe V_1 ou V_2 ainsi que séquence V_1V_2 correspondante) ont été extraites à l'aide d'un script Praat conçu par le second auteur. Ces données constituent des trajectoires temporelles (position temporelle, f_0, $F1$, $F2$, $F3$) associées à chaque séquence V_1V_2 qui sont ensuite traitées par un programme de manipulation et d'analyse des données dans l'environnement R (R Core Team, 2012). L'extraction de ces trajectoires permet de travailler à la fois sur les valeurs statiques de fréquences formantiques prises à différents instants dans la séquence mais également d'exploiter la composante dynamique en extrayant des valeurs de pente tout au long d'une transition (dérivée en chaque point de la trajectoire).

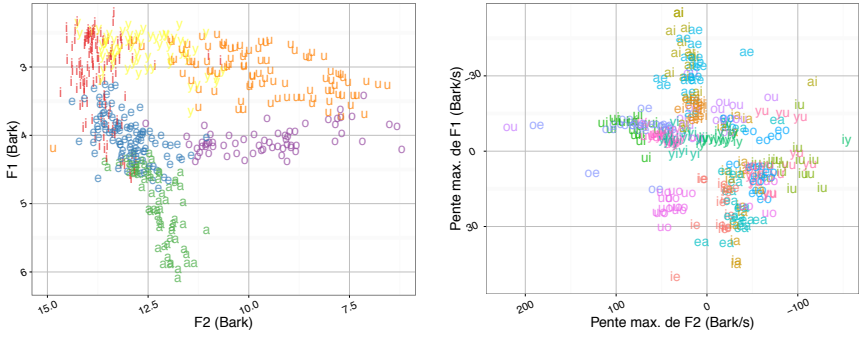
Les données de fréquences ont été transformées en échelle Bark puis normalisées en scores z (variable centrée réduite) pour chaque locuteur (normalisation de Lobanov). Les graphiques présentés dans cet article représentent les données en Bark (ou en Bark/s pour les pentes) afin de faciliter leur lecture. Les analyses statistiques décrites ont toutes été réalisées sur la base des valeurs normalisées, ce qui permet d'améliorer la discriminabilité entre classes dans les Analyses Discriminantes Linéaires. Afin de procéder au traitement des pentes maximales des trajectoires, celles-ci étaient préalablement lissées par une modélisation par courbes splines. La dérivée en chaque point de cette trajectoire était calculée puis la valeur maximale de cette dérivée dans la portion médiane de la séquence V_1V_2 (correspondant à la zone de transition) était extraite. Ceci nous permet d'obtenir une valeur de pente maximale correspondant à la vitesse de transition la plus élevée dans cet intervalle.

3. *Inter-Stimulus Interval* (Intervalle Inter-Stimulus)

3 Résultats

Afin de fournir une première approche des données acoustiques et de leur dispersion préalablement à l'analyse de classification, les fréquences formantiques mesurées au milieu de chaque voyelle et les valeurs de pentes maximales des séquences V_1V_2 produites par le locuteur 2 uniquement sont présentées dans la Fig. 1. La Fig. 1a est classique et permet de visualiser la répartition des réalisations vocaliques pour ce locuteur dans l'espace $F1 \sim F2$. La Fig. 1b représente les pentes maximales mesurées pour chaque séquence réalisée par ce locuteur. On notera que les données globales (les données de l'ensemble des 5 locuteurs) ne changent pas considérablement ces observations subjectives. Les espaces de fréquence des formants sont toujours nettement moins « confus » que ceux des pentes maximales. Les coefficients de variation (écart-type divisé par moyenne) associés aux différentes mesures semblent confirmer cette observation des espaces de variation (Tab. 2).

S'il ressort assez nettement que les données de pente mesurées sont plus difficiles à segmenter en catégories que les données de fréquence des formants, il faut nuancer cette première impression pour deux raisons. Le graphique des fréquences (Fig. 1a) n'est consitué que de 6 classes alors que celui des pentes maximales (Fig. 1b) est composé de 16 catégories. Il est donc prévisible qu'il devrait être plus difficile d'identifier des groupements cohérents « à variation équivalente ». Par ailleurs, si les données servant de base aux arguments de Carré (2009) étaient essentiellement fondées sur les mesures des deux premiers formants, une classification plus visible pourrait apparaître dans un espace multi-dimensionnel impliquant plusieurs mesures acoustiques.



(a) (b)

FIGURE 1 – Représentation (pour le locuteur 2 uniquement) dans un plan $F1 \sim F2$ (a) des valeurs de formants mesurées à 50% de la durée totale de la voyelle (en Bark) ; (b) des valeurs de pentes maximales (en Bark/s).

	Fréquence	*Pente maximale*
F1	*0.16 (0.03)*	*1.71 (1.68)*
F2	*0.12 (0.05)*	*5.43 (20.69)*
F3	*0.04 (0.01)*	*8.75 (18.36)*

TABLE 2 – Moyennes (et écarts-types) des coefficients de variation (écart-type divisé par moyenne) associés à chaque classe (V pour les fréquences, VV pour les pentes).

3.1 Analyses Discriminantes Linéaires

Afin d'approfondir notre compréhension de cette question, et d'explorer plus en détails la piste d'une discriminabilité des classes fondée sur les mesures de vitesse de transition qui pourrait reposer sur un espace multi-dimensionnel, nous avons donc conduit une série d'analyses discriminantes linéaires. Ces analyses pourraient nous permettre de faire ressortir une organisation des données acoustiques qui peut ne pas apparaître dans des espaces en 2 dimensions ou dans des comparaisons de moyennes / médianes et / ou de coefficients de variation.

Les mesures acoustiques de production ont été analysées à travers deux séries d'analyses discriminantes linéaires (LDA). Dans la première série (données statiques), les mesures de $F1$, $F2$, $F3$ et f_0 prélevées au milieu de chaque voyelle (V_1 et V_2) sont utilisées comme prédicteurs des 6 classes vocaliques impliquées. Dans la seconde série d'analyses, les mesures de pente maximale des transitions de $F1$, $F2$, $F3$ et f_0 entre 2 voyelles d'une séquence V_1V_2 sont utilisées comme prédicteurs des 16 catégories de séquences V_1V_2. Notre objectif est d'examiner dans quelle mesure les performances de classification reposant sur des données de pente maximale entre deux segments peuvent être plus « efficaces » que celles qui reposent sur les mesures de fréquence prises au milieu d'un segment.

3.1.1 Données statiques

Le tableau 3 présente les pourcentages de classification correcte des différentes classes vocaliques en fonction des trois groupes de prédicteurs utilisés. Cette analyse nous permet de voir que le pourcentage global de classification correcte varie peu avec l'ajout de prédicteurs. Il faut rentrer en détail dans les catégories vocaliques pour apercevoir un impact, comme celui de $F3$ sur le pourcentage de classification de /y/ pour atteindre un taux global de performance correcte de l'ordre de 70%.

	i	e	a	o	u	y	**global**
$F1 + F2$	78.0	68.2	73.3	66.2	76.5	22.6	**67.2**
$F1 + F2 + F3$	71.9	67.2	74.6	66.2	73.3	60.9	**69.5**
$F1 + F2 + F3 + f_0$	71.9	69.1	73.9	66.2	72.9	60.9	**69.8**

TABLE 3 – Pourcentages de classification correcte associés à chaque voyelle pour les groupes de prédicteurs $F1 + F2$, $F1 + F2 + F3$, $F1 + F2 + F3 + f_0$ (mesure prise au milieu de la voyelle).

.

Afin de pouvoir comparer des taux de performance issus de « tâches » dans lesquelles le nombre de catégories diffère, nous avons appliqué la formule utilisée par Schwartz *et al.* (2004). Cette formule de correction de la proportion de réponses correctes en fonction du nombre de catégories (Éq. 1) permet de représenter la taille relative de la différence entre proportion brute et proportion théorique en fonction du nombre de catégories disponibles.

$$\frac{p - p_{ref}}{1 - p_{ref}} \times 100 \quad (1)$$

Nous appellerons cette taille relative le « score corrigé ». Dans l'équation 1, p représente la proportion de réponses correctes mesurée et p_{ref} la proportion théorique de réponses au hasard ($\frac{1}{6}$ pour 6 catégories par exemple). Les scores corrigés pour le nombre de catégories sont présentés dans le tableau 4. Les données sont évidemment directement comparables au données du tableau 3. Elles seront principalement utiles pour la comparaison avec les données « dynamiques ».

	i	e	a	o	u	y	**global**
$F1+F2$	58.0	48.2	53.3	46.2	56.5	2.6	**44.1**
$F1+F2+F3$	51.9	47.2	53.6	46.2	53.3	40.9	**48.9**
$F1+F2+F3+f_0$	51.9	49.1	53.9	46.2	52.9	40.9	**49.2**

TABLE 4 – Mesures corrigées (Schwartz *et al.*, 2004) des pourcentages présentés dans la table 3.

3.1.2 Données dynamiques

Le but de cette section est d'évaluer la contribution des indices « dynamiques »de pente maximale présents dans les signaux de parole produits par ces locuteurs. Dans cette série d'analyses discriminantes, nous avons cherché à modéliser la classification en 16 classes V_1V_2 en prenant comme prédicteurs les pentes maximales extraites des transitions entre V_1 et V_2. Les mêmes mesures de $F1$, $F2$, $F3$ et f_0 ont servi de prédicteurs mais dans cette partie, ce sont les pentes maximales issues des trajectoires temporelles de ces indices qui ont été utilisées : les pentes maximales correspondent à la vitesse de changement la plus élevée au cours d'une transition. Les données de performance brutes et corrigées sont présentées respectivement dans les tableaux 5 et 6.

	/ie/	/ei/	/ia/	/ai/	/iu/	/ui/	/iy/	/yi/	
$F1+F2$	16.0	76.6	5.1	64.4	72.3	46.0	23.1	47.5	
$F1+F2+F3$	26.7	76.6	20.3	74.7	63.9	55.2	60.0	66.2	
$F1+F2+F3+f_0$	45.3	76.6	45.6	73.6	60.2	62.1	61.5	66.2	
	/ea/	/ae/	/eo/	/oe/	/ou/	/uo/	/uy/	/yu/	**global**
$F1+F2$	67.5	31.0	48.7	23.8	5.6	55.8	24.3	38.5	**41.0**
$F1+F2+F3$	70.1	47.6	46.1	44.0	16.7	48.1	51.4	50.0	**51.4**
$F1+F2+F3+f_0$	72.7	48.8	43.4	40.5	12.5	45.5	52.7	41.0	**53.3**

TABLE 5 – Pourcentages de classification correcte associés à chaque séquence VV pour les groupes de prédicteurs $F1+F2$, $F1+F2+F3$, $F1+F2+F3+f_0$ (pentes maximales de la transition).

.

Le pourcentage brut de classification correcte semble refléter des performances assez nettement inférieures à la classification atteinte à partir des mesures de fréquence statiques. Par ailleurs, on observe de manière similaire à ce qu'on observait sur les données statiques, la présence de certaines classes particulièrement mal catégorisées (/ie/, /ia/, /iy/) pour lesquelles l'ajout de $F3$ améliore parfois sensiblement la classification. La séquence /ou/ ne dépasse cependant jamais les 20% de classification correcte. Dans l'ensemble, le taux maximal de performance correcte brute atteint environ 53%, ce qui est nettement moins élevé que ce qu'on obtenait avec les mesures statiques (environ 70%).

Si l'on se penche sur les scores corrigés par contre, les différences entre les deux ensembles de données s'atténuent fortement. On parvenait à un score corrigé maximal de 49.2% sur la base de mesures de fréquence associées aux 6 catégories vocaliques alors qu'on atteint 46.4% sur la base des pentes maximales associées aux 16 classes de séquences V_1V_2. Ces deux valeurs semblent tout à fait comparables.

	/ie/	/ei/	/ia/	/ai/	/iu/	/ui/	/iy/	/yi/	
$F1+F2$	9.3	70.0	-1.6	57.7	65.6	39.3	16.4	40.8	
$F1+F2+F3$	20.0	70.0	13.6	68.0	57.2	48.5	53.3	59.6	
$F1+F2+F3+f_0$	38.7	70.0	38.9	66.9	53.6	55.4	54.9	59.6	
	/ea/	/ae/	/eo/	/oe/	/ou/	/uo/	/uy/	/yu/	**global**
$F1+F2$	60.9	24.3	42.0	17.1	-1.1	49.2	17.7	31.8	**33.7**
$F1+F2+F3$	63.5	41.0	39.4	37.4	10.0	41.4	44.7	43.3	**44.4**
$F1+F2+F3+f_0$	66.1	42.1	36.8	33.8	5.8	38.8	46.0	34.4	**46.4**

TABLE 6 – Mesures corrigées (Schwartz *et al.*, 2004) des pourcentages présentés dans la table 5.

4 Discussion

Des mesures de variation acoustique et des performances brutes issues des LDA , il ne ressort aucune tendance permettant d'affirmer la moins grande variabilité des pentes transitionnelles par rapport aux fréquences statiques. Au contraire même, les mesures de pente maximale semblent donner lieu à une plus grande variation univariée (cf. les mesures de coefficients de variation associées à $F1$, $F2$ et $F3$ et à leurs pentes maximales) ainsi qu'à des performances de classification correcte nettement moins élevées que celles obtenues à partir des mesures statiques. Cependant nous avons noté que le nombre de catégories possibles entre les LDA « statiques » et « dynamiques » diffère (6 voyelles $\rightarrow$ 16 paires V_1V_2). De cette augmentation du nombre de catégories résulte une diminution du pourcentage théorique de réponse au hasard, aussi le pourcentage théorique de réponse obtenu dans les tableaux 3 et 5, doit être observé par rapport au taux de réponse au hasard, ce qui est proposé à partir des scores corrigés dans les tableaux 4 et 6. Ces résultats ne nous permettent pas de conclure en faveur d'une hypothèse ou d'une autre même si l'écart de performance est très légèrement favorable à l'hypothèse « statique ».

Le prolongement de ces analyses se fera par l'ajout de $F4$ comme prédicteur. Ces mesures sont probablement cruciales pour la classification associée aux voyelles arrondies notamment et pourraient modifier considérablement les résultats des LDA. Nous étudierons aussi en détails les matrices de confusion obtenues afin d'analyser plus précisément la structure des réponses de classification qui pourrait être masquée par les mesures de performance globale. Les méthodes de *resampling* (*bootstrap / permutation*) permettront d'évaluer les intervalles de confiance des mesures de performance observées. Il est par ailleurs délicat d'interpréter une comparaison de résultats de classification impliquant deux ensembles différents de prédicteurs. Nous prévoyons une adaptation de cette procédure qui permettrait d'intégrer dans la même analyse l'ensemble des prédicteurs, ce qui rendrait possible le classement des prédicteurs entre eux.

Ces résultats devront aussi être confrontés aux données perceptives de Divenyi (2009) qui semblent confirmer la prise en compte de la vélocité des transitions dans la perception de séquences de voyelles synthétiques par les locuteurs. Enfin, nous explorerons le traitement de la composante temporelle dans les modélisations par systèmes dynamiques. En effet, dans les méthodes FDA (Functional Data Analysis) une abstraction du temps est nécessaire (Lancia & Tiede, 2012), cela implique donc la suppression de la vitesse or, selon l'hypothèse de Carré (2009), la vitesse serait l'élément essentiel car le plus stable, ce qui semble entrer en contradiction avec la notion d'abstraction temporelle.

Références

ANDRUSKI J. E. & NEAREY T. M. (1992). On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables. *The Journal of the Acoustical Society of America*, **91**, 390–410.

BOERSMA P. & WEENING D. (2014). Praat : Doing phonetics by computer. Computer program. Version 5.4.

CARRÉ R. (2009). Signal dynamics in the production and perception of vowels. In F. PELLEGRINO, E. MARSICO, I. CHITORAN & C. COUPÉ, Eds., *Approaches to Phonological Complexity*, p. 59–81. Berlin – New-York : Mouton de Gruyter.

CONTENT A., MOUSTY P. & RADEAU M. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, **90**(4), 551–566.

DIVENYI P. (2009). Perception of complete and incomplete formant transitions in vowels. *The Journal of the Acoustical Society of America*, **126**(3), 1427–1439.

DURAND J. (2005). Les primitives phonologiques : des traits distinctifs aux éléments. In N. NGUYEN, S. WAUQUIER-GRAVELINES & J. DURAND, Eds., *Phonologie et Phonétique : Forme et Substance*, chapitre 3, p. 63–93. Paris : Hermès.

HILLENBRAND J. M., CLARK M. J. & NEAREY T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, **109**(2), 748–763.

JENKINS J. J., STRANGE W. & TRENT S. A. (1999). Context-independent dynamic information for the perception of coarticulated vowels. *The Journal of the Acoustical Society of America*, **106**(1), 438–448.

LANCIA L. & TIEDE M. (2012). A survey of methods for the analysis of the temporal evolution of speech articulator trajectories. In S. FUCHS, M. WEIRICH, D. PAPE & P. PERRIER, Eds., *Speech Planning and Dynamics*, p. 239–271. Frankfurt-am-Main, Germany : Peter Lang.

LINDBLOM B. & STUDDERT-KENNEDY M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical society of America*, **42**(4), 830–843.

NEAREY T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, **85**(5), 2088–2113.

O'SHAUGHNESSY D. (1986). The effects of speaking rate on formant transitions in French synthesis-by-rule. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, volume 11, p. 2027–2030.

PETERSON G. & BARNEY H. (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America*, **24**(2), 175–184.

R CORE TEAM (2012). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing. ISBN 3-900051-07-0.

SCHWARTZ J. L., BERTHOMMIER F. & SAVARIAUX C. (2004). Seeing to hear better : evidence for early audio-visual interactions in speech identification. *Cognition*, **93**, B69–B78.

STRANGE W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*, **85**(5), 2081–2087.

STRANGE W., EDMAN T. & JENKINS J. (1979). Acoustic and phonological factors in vowel perception. *Journal of Experimental Psychology : Human Perception and Performance*, **5**, 643–656.

STRANGE W., JENKINS J. & JOHNSON T. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, **74**(3), 695–705.